Checking your imagination: applications of the free R value

Gerard J Kleywegt[1] and Axel T Brünger[2]*

Addresses: [1]Department of Molecular Biology, Biomedical Centre, Uppsala University, Box 590, S-751 24 Uppsala, Sweden and [2]The Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA.

*Corresponding author

E-mail: brunger@laplace.csb.yale.edu

Structure 15 August 1996, 4:897–904

 ISSN 0969-2126

Introduction

In 1990, Brändén and Jones [1] suggested that the process of electron-density map interpretation involves a degree of subjectivity on the part of the crystallographer. In the worst case, serious errors may be introduced into a model. In particular, their commentary stimulated the development of many different approaches for assessing the quality of protein crystal structures. Many of these methods validate models using statistical surveys of high-resolution X-ray structures (either of small molecules or of macromolecules) from which expected values or ranges of values are calculated for structural features. However, without the use of experimental diffraction data, one can never be sure whether a statistical 'outlier' (for example, in a Ramachandran plot) represents an error in the model or a genuine feature of the crystal structure [2], and the distinction between the two is obviously important. The ambiguity can only be resolved if the diffraction data is used in the quality assessment. Therefore, methods to assess if the model is an accurate representation of the experimental data must rely on statistics that involve both the diffraction data and the atomic model. Furthermore, few statistical surveys have been performed for macromolecular structures other than proteins, such as RNA, which means that knowledge-based rules are not available for the latter.

The quality of the fit of a model to the diffraction data is given by the R value, which measures the discrepancy between the observed (F_o) and calculated (F_c) structure-factor amplitudes:

$$R=\Sigma| |F_o|-|F_c| |/\Sigma|F_o|$$

This statistic is closely related to the crystallographic residual, $\Sigma(|F_o|-|F_c|)^2$, which is minimized during structure refinement. However, this R value can be made arbitrarily low by increasing the number of adjustable parameters used to describe the model. Therefore, a low R value does not necessarily correspond to a good model. Indeed, several cases are known in which models with serious errors were refined to 'acceptable' R values [1,3]. To overcome this problem, the method of statistical cross-validation was introduced into X-ray crystallography [3–5]. In this scheme, the diffraction data are divided into two sets: a large 'working' set (usually comprising 90% of the data), and a complementary 'test' set (comprising the remaining 10%). The data in the working set are used in the normal crystallographic refinement process, whereas the test data are not. At any stage of the refinement, an R value may be calculated for the test set, which is referred to as the free R value, or R_{free}. The free R value measures the degree to which the model predicts the diffraction data for the test set which was excluded from the modeling and refinement process. The nature of crystallographic diffraction data is such that every reflection contains information about the entire structure. Therefore, changes to a model that do not improve the model's ability to describe the diffraction data should not improve the fit of the model to the test set, and thus, the free R value would remain constant or increase [3–5]. The free R value has been shown to be correlated with the accuracy of atomic models. In practice, this means that models with serious errors can be identified by a very high free R value (>0.40) irrespective of the value of the conventional R value, which may be very low (~0.20). Furthermore, the free R value is a statistic for assessing the improvement of a model during the course of refinement and rebuilding. This makes it possible to formulate alternative hypotheses about the model, and to test the validity of these hypotheses by inspection of the free R value. Such hypotheses may pertain to the way in which temperature factors, non-crystallographic symmetry (NCS), bulk solvent or conformational flexibility are modeled. Alternatively, different refinement protocols or strategies can be tested and their performance compared using the free R value. An additional advantage of the free R value over knowledge-based validation methods is that it can be applied to any type of model and does not depend on the availability of database-derived knowledge. It should also be noted that cross-validation can be applied to any statistic, such as likelihood [6,7] or phase error [8].

Here we review applications of the free R value, and discuss practical issues and caveats related to the use and interpretation of this statistic. We also present a survey of free R values of published X-ray crystal structures of macromolecules.

Applications

Detection of errors

The usefulness of the free R value was initially demonstrated by the empirical observation that it is highly

correlated with the phase error of a model [3]. Examples of applications of the free R value included the optimization of weights (see section below), and the detection of errors in crystal structures, such as the partial mistracing of the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco) [3,9,10]. Whereas the correct rubisco model had a free R value of 0.34, the incorrect one had a value of 0.47. Another example was given by intentionally tracing backwards the structure of cellular retinoic-acid-binding protein type II (CRABP II), previously solved at 1.8 Å resolution [11] and refining this model using data to only 3 Å resolution [12,13]. Using standard refinement methods and procedures, the conventional R value decreased to 0.214 with 'excellent stereochemistry.' The free R value, on the other hand, converged to a value of 0.617, slightly worse than the value expected for a random set of scatterers [14]. These examples demonstrate that, especially when the ratio of the number of diffraction observations to model parameters is low, cross-validation is necessary even though it requires that a small subset of the reflections is excluded from the refinement. At high resolution (better than 2.0 Å), the conventional R value can sometimes be used as an indicator of model quality. For instance, when the backward-traced CRABP II model was refined to 1.8 Å resolution, the conventional R value remained above ~0.35. However, even at high resolution the free R value is useful to avoid over-fitting (see sections below).

The free R value can sometimes identify problems with the diffraction data. For example, the free R value remained at 0.35 during the refinement of the structure of holo CRABP type I [11] against a data set that had been processed to 2.5 Å resolution. Re-examination of the original image plate data showed that the resolution limit of the data had been grossly over-estimated. More careful reprocessing yielded a data set with a nominal resolution of only 2.9 Å, with relatively weak and incomplete data in the highest shells (effective resolution ~3.2 Å). Refinement of the model against the reprocessed data set produced an R value of 0.251 and a free R value of 0.320 [11]. The free R value indicated the poor quality of the original 2.5 Å data in the higher resolution shells; beyond ~3.2 Å the free R value increased sharply to levels of 0.45–0.55, indicating that any information present in the diffraction data in these shells was buried in the noise.

Optimization of refinement protocols

At low resolution, high temperature simulated annealing [15] may not necessarily improve a model (unless, for instance, phase restraints are used). The structure of the complex between the Fc fragment of human immunoglobulin IgG and the C2 domain of protein G was refined against a data set with an effective resolution of ~3.5 Å [16]. Using the free R value as a guide, it was found in this case that none of the many simulated annealing protocols used yielded an improved model; the free R value remained constant or even increased, even though the conventional R value often dropped by 0.10. This is not a general rule, however, and the applicability of simulated annealing refinement should be investigated by inspection of the behavior of the free R value on a case by case basis. For example, initial attempts to apply simulated annealing refinement to a rough homology model failed in the case of *Trichoderma Reesei* endoglucanase I (EGI) (GJK *et al.*, unpublished data) at 4.0 Å (the free R value remained above 0.50). A $2F_o–F_c$ map was calculated using a polyalanine model of one of the probe molecules. This map was poor, but after 15 cycles of twofold NCS averaging a significantly improved map was obtained. Using this map, ~75% of the sequence could be assigned to the model, yielding a starting value of ~0.45 for both the R and free R values. After a 4000 K slow-cooling refinement with positional restraints for the Cα atoms, the free R value dropped to 0.39 (and the R value to 0.28). In the resulting averaged map another ~15% of the model (60 residues) could be traced and built, indicating that the simulated annealing refinement had significantly improved the model, even at 4.0 Å resolution.

The free R value can also be used to optimize the relative weights of the various contributions to the empirical energy function used in refinement [3,4]. The free R value was used to obtain the optimal set of weights for the Engh and Huber parameters [17]. As the original Engh and Huber parameters effectively increased the weights for bond lengths and bond angles, the dihedral and improper torsion angles were underweighted. This was shown clearly by the behavior of the free R value [4] and led to a readjustment of the weights for these terms.

The free R value can also be used to determine the best weight on the crystallographic pseudo-energy term(s), such as the residual between the observed and calculated structure-factor amplitudes [4], and to optimize the weight of temperature factor restraints relative to the X-ray pseudo-energy term [3]. The weight for the crystallographic pseudo-energy term as estimated by version 3.1 of X-PLOR [18] tends to be too high (i.e., it weighs the X-ray term too heavily, which may lead to over-fitting and poorer geometry). Running identical refinements using different values for this weight can be used to find the optimal weight on a case by case basis. The weight that yields the lowest free R value can then be used in subsequent refinement rounds. It is of interest to note that this procedure may lead to tightly restrained geometry, tighter even than observed for (atomic resolution) small-molecule structures. However, it is important to realize that this tight geometry is a consequence of the information content of the crystallographic diffraction data [4]. Apparently typical macromolecular X-ray data does not contain sufficient information to

produce similar geometric distributions around the mean values as those observed for atomic resolution small-molecule structures.

Prevention of over-fitting

The free R value can be used to judge if changes to a model improve the model's accuracy or lead to over-fitting. It can also be employed to assess the validity of alternative methods to model thermal motion, disorder, or NCS. The validity of these models depends on the quality and completeness of the diffraction data, a highly redundant 2.8 Å synchrotron data set may enable modeling of isotropic atomic temperature factors, whereas a poor or incomplete 2.4 Å data set may not. Similarly, the quality of the starting model is important (e.g., a 1.5 Å model obtained by difference-map techniques is probably better than a 3.0 Å multiple isomorphous replacement [MIR] model).

To assess if refinement of individual temperature factors (B factors) is warranted, refinement should be carried out both with residue-grouped and restrained individual temperature factors. If the model with individual temperature factors does not have a significantly lower free R value, one can conclude that with the current model and the present data set, temperature factors are best modeled by group.

NCS can be treated in different ways during refinement. If all copies of a molecule are forced to be identical, and the NCS operators that relate the individual molecules are kept fixed during a round of refinement, the NCS is said to be constrained. If small differences between the molecules are allowed, and the NCS operators are allowed to change during refinement, the NCS is said to be restrained. At high resolution one may be able to refine NCS-related molecules independently [12,19]. One can test if replacing NCS constraints by restraints yields a significantly better model for the data; an example of this using data from A2U is shown in Table 1. Clearly, the free R value indicates that the diffraction data for A2U are best modeled by forcing the four monomers to be identical or very similar. If no NCS restraints are used at all, the root mean square (rms) difference between the monomers is 0.87 Å, a significant departure from the best model predicted by the free R value. Although this model has excellent stereochemistry, it requires four times as many adjustable parameters and leads to an increase in the free R value, suggesting over-fitting; the unrestrained model is not the most faithful description of the diffraction data of A2U.

Another example of the use of the free R value to determine the best NCS model is in the refinement of the GroEL structure at 2.8 Å resolution [20]. The crystal structure of GroEL showed the protein to contain three domains, with seven copies of the molecule present in the asymmetric unit. Due to differences in the relative orientations of the three domains in each of the seven copies of the molecule, NCS restraints were more appropriate in this case than NCS constraints. As the three domains had very different average temperature factors, the NCS restraint weights were set differently for each of them, and these values were optimized using the free R value. At the end of the refinement, the best results were obtained with high weights on the NCS restraints. This indicated that, although the domains have different orientations in the seven molecules, their local structures are very similar given the information content of the diffraction data.

Even in refinements at atomic resolution, the free R value can be used to prevent over-fitting. For instance, Sheldrick and co-workers use the free R value to assess if the introduction of anisotropic temperature factors and hydrogen atoms (either explicitly defined, or 'riding' on heavier atoms) improves the phase accuracy of a model [21].

Coordinate-error estimates

As the conventional R value shows little correlation with the accuracy of a model (unless the observable : parameter ratio is high), coordinate-error estimates derived from Luzzati [22] or σ_A plots [23] are unrealistically low. It was therefore suggested that more reliable coordinate errors may be estimated from a cross-validated Luzzati [11] or σ_A plot [5] instead.

Figure 1 shows the results of calculations using the crystal structure and diffraction data of the enzyme penicillopepsin [24]. At 1.8 Å resolution, the model has an estimated coordinate error of ~0.2 Å as assessed by multiple independent refinements. As the resolution is lowered and additional refinement carried out, the coordinate error increases monotonically. However, the conventional R value actually improves as the resolution gets lower and the quality of the model decreases. Consequently, estimates of the coordinate error obtained from Luzzati [22] or σ_A plots [23] do not display the correct behavior either; the error estimates are approximately constant regardless of the resolution and actual coordinate error of the models. However, when cross-validation is used (i.e., the test reflections are used to compute the estimated coordinate errors [5,11]) the results are much better; the cross-validated errors are close to the actual rms distances to the original crystal structure, and they show the correct trend as a function of resolution.

In the case of the backward-traced CRABP II model, the estimated coordinate error (based on a Luzzati plot using the conventional R value) is ~0.35 Å, whereas that based on the free R value is 'infinite', which is more appropriate in this case.

Phase-error estimates

Although the free R value is highly correlated with the mean absolute phase error, there is no simple relationship between the two that enables one to determine the

Table 1

Tests of various NCS models at low resolution.*

NCS weight (kcal mol^{-1} Å^{-2})	R_{cryst}	R_{free}	$\sigma(R_{free})^{\dagger}$	R_{free} range	R_{free}–R	Rmsd(Cα)‡	Rmsd(all)§
'infinity'	0.242	0.262	0.005	0.256 – 0.271	0.020	0	0
300	0.235	0.262	0.005	0.257 – 0.270	0.027	0.02	0.04
100	0.229	0.263	0.004	0.258 – 0.270	0.034	0.05	0.08
75	0.227	0.264	0.004	0.260 – 0.272	0.037	0.06	0.09
25	0.220	0.268	0.004	0.262 – 0.273	0.048	0.12	0.17
10	0.214	0.272	0.004	0.267 – 0.278	0.058	0.18	0.26
5	0.211	0.275	0.005	0.269 – 0.282	0.064	0.23	0.35
'0'	0.206	0.286	0.005	0.280 – 0.294	0.080	0.43	0.87

*Results of using NCS constraints, restraints or no restraints with slow-cooling simulated annealing protocols (starting temperature 2 000 K; temperature steps of –25 K; followed by energy minimization) at 2.8 Å resolution. The final 2.5 Å model of A2U (GJK *et al*., unpublished data), was stripped of water molecules and used as the starting model (initial R and R_{free} 0.25). For each NCS protocol, complete cross-validation was carried out by performing ten runs with non-intersecting 10% test sets. The refinements with an NCS force constant 'infinity' used strict NCS constraints; those with a force constant 'zero' used no NCS restraints and yielded the poorest model. All models had good stereochemistry (rms deviations from ideality of < 0.01 Å for the bond lengths and < 1.4° for the bond angles). †Standard deviation of the free R value around the average for the different test sets. ‡Average rms deviations on Cα atoms between the four monomers. §Average rms deviations for all non-hydrogen atoms.

absolute phase error from the free R value. Lunin and Skovoroda [8] have developed a method to estimate the magnitude of the phase error of a partial atomic model that contains errors. When the method is applied to the working set (diffraction data used in the refinement) for a simulated test case, the estimated phase errors are considerably lower than the actual ones. In contrast, when the test set of reflections is used for the model phase-error calculation, the resulting phase-error estimates are very close to the actual values. Thus, cross-validation allows the computation of more realistic phase-error estimates.

Descriptions of disorder and solvation

Cross-validation has also been used in efforts to improve the description of crystal structures. Conformational variability in penicillopepsin was modeled using a multiple-conformer approach [25]. Cross-validation showed that this technique yielded better phase accuracy than either single-conformer models or time-averaged molecular dynamics. In this particular case, an ensemble of eight copies of the molecule yielded the best description of the diffraction data.

Several methods of modeling bulk solvent in protein crystals have been compared using complete cross-validation [26]. In the particular cases studied, it was concluded that a simple flat solvent model represents a fairly accurate description of the diffraction data, and that more sophisticated models only produce marginal improvements. However, each case is different and the free R value can be used to decide which model is optimal given the quality and information content of the diffraction data.

Density modification

Cross-validation has also been used to validate and optimize density-modification procedures. The interested reader is referred to the papers by Baker *et al*. [27], Grimes and Stuart [28], Roberts and Brünger [29], and Cowtan and Main [30].

Practical considerations

Independence of the test set: correlated reflections

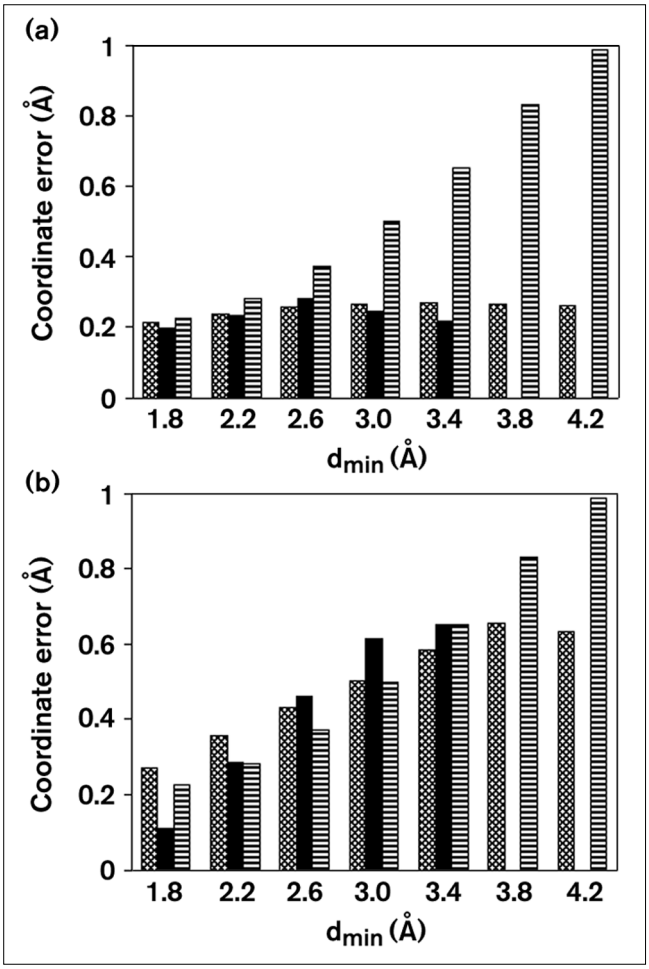
The basic assumption upon which the free R value is founded is that the reflections in the test set are not correlated with any of those in the working set. For macromolecular crystallography, however, such relationships can exist. Due to the contrast between solvent and protein, every reflection is related through the G-function [31] to some of its neighbors in reciprocal space. No method of selecting the test set of reflections will completely exclude these relationships. On the other hand, as the case of the backward-traced CRABP II model shows [12,13], in practice the effect appears to be small. Nevertheless, there are other cases in which the relationships are much stronger, even to such an extent that the test reflections become noticeably biased.

Serious correlations between reflections are introduced when a crystallographic symmetry operator is incorrectly modeled as an NCS operator. This type of space group error cannot be detected by the free R value, as most of the test set reflections will have a symmetry-equivalent in the working set [32].

A similar situation may arise when anomalous data has been collected and is used in refinement without appropriate precautions. In this case, it is imperative that Friedel pairs are either both in the test set or both in the working set; otherwise many of the test set reflections will be highly coupled to the working set.

Complications may also occur when new, isomorphous diffraction data is obtained (either for the same molecule, or

Figure 1



The effect of resolution on coordinate-error estimates: accuracy as a function of resolution. Refinements were begun with the crystal structure of penicillopepsin with water molecules omitted and with uniform temperature factors. The low-resolution limit was set to 6 Å. The penicillopepsin diffraction data were artificially truncated to the specified high-resolution limit (d_{min}). Each refinement consisted of simulated annealing using a Cartesian space slow cooling protocol starting at 2000 K, overall B-factor refinement, and individual restrained B-factor refinement. All refinements were carried out with 10% of the diffraction data randomly omitted for cross-validation. (a) Coordinate-error estimates of the refined structures using Luzzati [22] (cross-hatched bar) and σ_A plots (black bar) [23]. All observed diffraction data were used (i.e., no cross-validation was performed). The actual coordinate errors (rms differences to the original crystal structure) are shown for comparison (bar shaded with horizontal lines). No σ_A estimates are shown below 3.4 Å resolution because the method became numerically unstable. (b) Cross-validated coordinate-error estimates. The test set was used to compute the coordinate-error estimates. Complete cross-validation [26] was used to compute σ_A (black bar) whereas averaging over individual error estimates for different test sets was used to compute Luzzati's cross-validated error estimate (cross-hatched bar). The actual coordinate errors are shown as a bar with horizontal lines.

for a complex or mutant). If the two data sets merge well, the new data does not provide an independent check on the quality of the model. In addition, if one starts refinement of an isomorphous complex or mutant one should either maintain the same set of test reflections, that were used in the refinement of the starting model, or perform high temperature simulated annealing to uncouple the working and free R values [4].

In the special case where there are two or more non-isomorphous crystal forms of the same molecule, the problems associated with relationships between reflections could be overcome by refining the model against the data for one crystal form, and calculating unbiased free R values using the data of one or more of the other crystal forms. This method, which has not been used yet to our knowledge, might be especially useful in the early stages of model building and refinement of a new structure to ensure correct chain tracing. The 'free' crystal form(s) might be complexes or mutants that are expected to have the same overall fold.

In the case of NCS, relationships between reflections exist that are not exact but are determined by the G-function [31]. The presence of NCS may in general reduce the difference between the R and free R values. For example, the structure of the MS2–RNA complex, with tenfold NCS, has an R value of 0.192 and a free R value of 0.209 [33]. To investigate if the relationships between the reflections could be strong enough to make gross tracing errors undetectable, test calculations have been carried out with A2U as this structure has fourfold NCS. The structure was intentionally traced backwards and refined against 6–3 Å data using different NCS models and different ways to select the test set reflections (see below); the results are shown in Table 2. Although the free R value did not reach the level it attained for the backward-traced CRABP II structure (0.62), it was never lower than 0.46, with constrained NCS (the maximum value was 0.55 with unrestrained NCS). These values are largely independent of the test set selection. Moreover, if the NCS was constrained or restrained, even the conventional R value could not be reduced to a satisfactory level (~0.35), whereas if the NCS was not restrained the R value easily dropped to ~0.27. Inclusion of water molecules and further refinement probably would have reduced the normal R value even further. This example, and other unpublished observations, suggest that use of the free R value is valid in the presence of NCS. In general, the absolute free R value for a structure with NCS will be lower than that for a similar structure without NCS, which may make comparisons of free R values difficult for structures with and without NCS. However, the free R value can always be used to monitor the progress of refinement of a particular crystal structure, regardless of the presence of NCS.

Standard deviation of the free R value

It was shown that the standard deviation of the free R value for different test sets depends mainly on the size of the test

set [5]. The relative error in the free R value appears to be inversely proportional to the square root of the number of test reflections [5,34].

$$\sigma(R_{free})/R_{free} \sim 1/\sqrt{N_{test}}$$

These standard deviations pertain to the variation of the free R value when cross-validation with different test sets of reflections is performed. However, a drop in the free R value smaller than $\sigma(R_{free})$ may still be meaningful if the same test set is used. If in doubt one could carry out several cross-validations with different test sets and determine if the free R value is reduced for all test sets [4].

Test set size and selection

It was suggested to use 10% of a unique set of reflections as the test data set [3]. However, because the standard deviation of the free R value computed for different test sets is inversely proportional to the square root of the number of test reflections, a minimum number of ~500 test reflections is recommended [5]. The average free R value (averaged over multiple test sets) appears to be rather insensitive to the size of the test set in the cases studied [4,5].

Different ways to select the test set have been devised; X-PLOR [35] assigns random reflections to the test set and SHELXL [21] sets aside every tenth reflection. In addition a program to generate test sets in thin resolution shells, or in small spheres in reciprocal space is available [36]. However, in the case studied, the free R value is rather insensitive to the selection procedure even when NCS is present (Table 2).

It is desirable to include all diffraction data when calculating electron-density maps to avoid truncation errors. In principle, the use of real-space refinement techniques [37] could introduce some bias towards the test reflections, but the seriousness of this effect has not been demonstrated. Moreover, if simulated annealing is used throughout the refinement, any model bias is likely to be removed during subsequent refinement.

When the refinement is complete, a final round of positional and temperature factor refinement can be carried out using all reflections (although the effect on the coordinates and temperature factors will usually be small). Nevertheless, it is desirable to quote the last recorded free R value in publications.

Acceptable values

It is difficult to give guidelines as to what constitutes an acceptable free R value, as every case is different. However, all cases involving grossly incorrect models have free R values in excess of 0.40. This means that free R values greater than 0.40 should be a reason for concern.

It is important to realize that this threshold only pertains to the final, refined model. For example, when a model is still very incomplete, one often observes relatively high free R values because only a fraction of the scattering matter is accounted for by the model. Furthermore, if a model is mainly correct but contains a few very large errors, the free R value may remain high.

Ideally, the difference between the crystallographic R value and free R value should be as small as possible. The difference between these two statistics depends on a number of factors [4]. Firstly, the quality of the data (e.g., as measured by R_{sym}, $\langle I/\sigma(I)\rangle$, completeness and multiplicity); the poorer the data the more likely it is that noise will be fitted, which will increase the discrepancy between R and R_{free}. Secondly, the global correctness of the model: clearly, if the model is grossly mistraced the free R value will remain very high, whereas the conventional R value may have an 'acceptable' value. Thirdly, the completeness of the model [4]. Finally, the degree of over-fitting; the smaller the ratio of diffraction observations to model parameters, the more a model will be over-fitted [12].

In order to obtain empirical information about the distribution of free R values, we analyzed the May 1996 release of the Protein Data Bank (PDB) [38] and performed a survey of some of the major journals in which macromolecular X-ray structures were published during the past six months. Of 3657 macromolecular X-ray structures in the PDB, 178 had an entry for the free R value (an increase from 62 entries in May 1995, and 95 in October 1995). The journal survey yielded 179 structures for which a free R value had been reported (Table 3). For 44% of the published structures a free R value is reported. However, this number varies widely depending on the journal.

Table 2

Effect of NCS on R and free R values for incorrect models.*

NCS	Constrained		Restrained		Unrestrained	
	R_{cryst}	R_{free}	R_{cryst}	R_{free}	R_{cryst}	R_{free}
Test set						
Random	0.365	0.465	0.347	0.484	0.268	0.522
Thin shells	0.360	0.477	0.348	0.485	0.266	0.552
Thick shells	0.361	0.470	0.351	0.531	0.267	0.531

*Listed are R and R_{free} values for test calculations with an intentionally backward-traced model of A2U (GJK *et al*., unpublished data). The calculations were performed to assess the effect of relationships between reflections in the case of NCS on the free R value for grossly incorrect models. The structure was subjected to simulated-annealing refinement using data with $F > 2\sigma(F)$ between 6.0 and 3.0 Å. Three different methods of selecting 10% test reflections were tried: random; in 15 thin resolution shells; and in 5 thick resolution shells. For each set of reflections three different NCS models were tested: constrained (observable : parameter ratio ~2.5); restrained; and unrestrained (observable : parameter ratio ~0.6).

Of the 357 structures included in our survey, only one had a free R value exceeding 0.40; the lowest recorded free R value was 0.154. The average R value was 0.20 (σ=0.03), and the average free R value 0.26 (σ=0.04). The average difference between the free R and R values is 0.07 (σ=0.04), but the range is very large (from 0.003 up to 0.176). The collected R and free R values, as well as the difference between the two, are shown in Figure 2 as a function of resolution. The free R value shows a slightly higher correlation with resolution than the normal R value (linear correlation coefficients of +0.52 and +0.47, respectively). The difference between the free R and R values is hardly correlated with the normal R value (correlation coefficient +0.12), but highly correlated with the free R value (+0.77).

Global versus local quality

As the free R value is a global statistic, it is not very sensitive to small changes in a model [4]. For instance, although the addition of the first 50 water molecules to a model may yield a sizable drop in the free R value, the addition of small numbers of water molecules will not affect the value noticeably. Similarly, minor local rebuilding which moves only a few atoms is not likely to have a major effect on the free R value. In all these cases other criteria have to be used, such as temperature factors, fit to the (omit) density (e.g., real-space R values), and comparison of the model to high-resolution database structures [39].

A posteriori *calculations*

If one wants to assess the accuracy of structures refined without cross-validation, *a posteriori* free R values appear

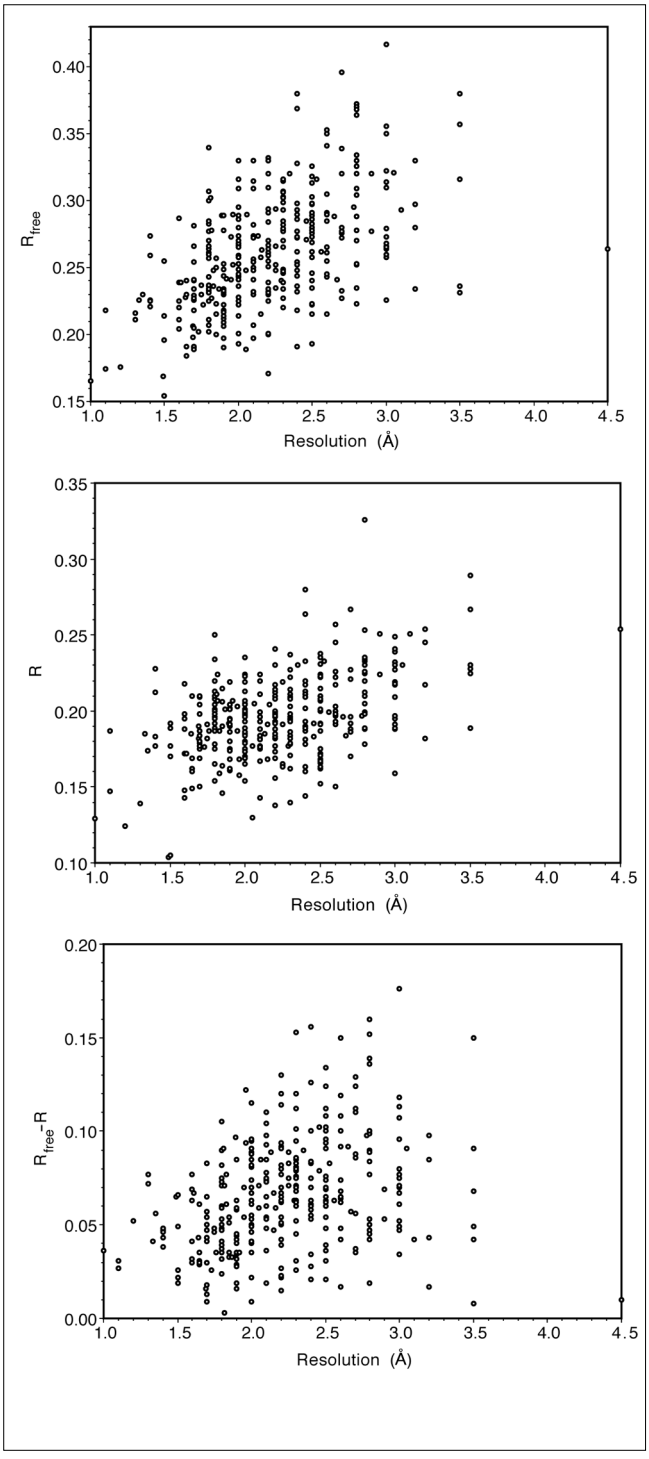
Table 3

Survey of reported free R values.*

Journal	Papers	Structures	R_{free}	% R_{free}
Nature	19	21	20	95
Cell	12	16	14	88
Structure	35	53	43	81
Science	10	15	10	67
Nat. Struct. Biol.	28	32	20	63
Acta Cryst. D	32	41	13	32
J. Mol. Biol.	54	92	25	27
Biochemistry	49	99	26	26
Protein Sci.	17	32	7	22
Proteins	4	5	1	20
TOTAL	260	406	179	44

*All issues of the listed journals that appeared between January 1 and June 30, 1996 were scanned manually for reports of macromolecular X-ray crystal structures. The number of such papers (column 'Papers'), the number of reported structures (column 'Structures') and the number of structures for which a free R value was quoted (column 'R_{free}') were tallied; the percentage of structures for which a free R value was quoted was computed from these numbers (column '% R_{free}'). In a few cases, the report indicated that the free R value had been used during refinement but no values were quoted.

Figure 2



Analysis of a survey of crystallographic R and free R values reported for macromolecular structures. The data were combined from an analysis of the May 1996 release of the PDB, and a survey of several journals (see the text and Table 3 for details). (a) Plot of free R values for 357 structures as a function of the resolution of the study. (b) Plot of the crystallographic R values, for the same structures as in (a), as a function of resolution. (c) Plot of the difference between free R and crystallographic R values as a function of resolution.

to reproduce the value that would have been obtained if the free R value had been used throughout, provided the simulated annealing refinement is started from sufficiently high temperatures. For example, when the backward-traced CRABP II structure was refined against all data, and then subjected to *a posteriori* calculation of the free R value, a starting temperature of 500 K yielded a free R value of 0.45. Thus, the simulated annealing refinement had to be started at 4000 K in order to reproduce the more realistic free R value of 0.62 [12,13].

Acknowledgements

GJK thanks Dr T Alwyn Jones for many stimulating discussions concerning model quality, validation and the free R value, and Dr Randy J Read for suggesting the test with a backward-traced model containing NCS. ATB thanks Drs P Adams, FT Burling, P Gros, J-S Jiang, RJ Read, LM Rice, and ALU Roberts for stimulating discussions about cross-validation in crystallography. GJK acknowledges support from Uppsala University and the European Union (through grant number BIO4-CT96-0189 to Dr TA Jones, Uppsala). ATB acknowledges support from the National Science Foundation (grant number BIR-9514819).

References

1. Brändén, C.I. & Jones, T.A. (1990). Between objectivity and subjectivity. *Nature* **343**, 687–689.
2. Kleywegt, G.J. & Jones, T.A. (1996). Good model-building and refinement practice. Methods Enzymol., in press.
3. Brünger, A.T. (1992). Free R-value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475.
4. Brünger, A.T. (1993). Assessment of phase accuracy by cross-validation: the free R-value. Methods and applications. *Acta Cryst. D* **49**, 24–36.
5. Brünger, A.T. (1996). The free R-value: a more objective statistic for crystallography. *Methods Enzymol*. in press.
6. Bricogne, G. (1984). Maximum entropy and the foundations of direct methods. *Acta Cryst. A* **40**, 410–445.
7. Pannu, N.S. & Read, R.J. (1996). Improved structure refinement through maximum likelihood. *Acta Cryst. A* **52**, in press.
8. Lunin, V.Yu. & Skovoroda, T.P. (1995). R-free likelihood-based estimates of errors for phases calculated from atomic models. *Acta Cryst. A* **51**, 880–887.
9. Chapman, M.S., Suh, S.W., Curmi, P.M., Cascio, D., Smith, W.W. & Eisenberg, D. (1988). Tertiary structure of plant RuBisCO: domains and their contacts. *Science* **241**, 71–74.
10. Curmi, P.M., Schreuder, H., Cascio, D., Sweet, R.M. & Eisenberg, D. (1992). Crystal structure of the unactivated form of ribulose-1,5-bisphosphate carboxylase/oxygenase from tobacco refined at 2.0 Å resolution. *J. Biol. Chem.* **267**, 16980–16989.
11. Kleywegt, G.J., *et al.*, & Jones, T.A. (1994). Crystal structures of cellular retinoic acid binding proteins I and II in complex with all-*trans*-retinoic acid and a synthetic retinoid. *Structure* **2**, 1241–1258.
12. Kleywegt, G.J. & Jones, T.A. (1995). Where freedom is given, liberties are taken. *Structure* **3**, 535–540.
13. Kleywegt, G.J. & Jones, T.A. (1995). Braille for pugilists. In *Making the Most of Your Model*. (Hunter, W.N., Thornton, J.M. & Bailey, S., eds), pp. 11–24, SERC Daresbury Laboratory, Warrington, UK.
14. Wilson, A.J.C. (1950). Largest likely values for the reliability index. *Acta Cryst*. **3**, 397–398.
15. Brünger, A.T., Krukowski, A. & Erickson, J.W. (1990). Slow-cooling protocols for crystallographic refinement by simulated annealing. *Acta Cryst. A* **46**, 585–593.
16. Sauer-Eriksson, A.E., Kleywegt, G.J., Uhlén, M. & Jones, T.A. (1995). Crystal structure of the C2 fragment of streptococcal protein G in complex with the Fc domain of human IgG. *Structure* **3**, 265–278.
17. Engh, R.A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Cryst. A* **47**, 392–400.
18. Brünger, A.T., Karplus, M. & Petsko, G.A. (1989). Crystallographic refinement by simulated annealing: application to a 1.5 Å resolution structure of crambin. *Acta Cryst. A* **45**, 50–61.
19. Kleywegt, G.J. (1996). Use of non-crystallographic symmetry in protein structure refinement. *Acta Cryst. D* **52**, 842–857.
20. Braig, K., Adams, P.D. & Brünger, A.T. (1995). Conformational variability in the refined crystal structure of the chaperonin GroEL at 2.8 Å resolution. *Nat. Struct. Biol.* **2**, 1083–1094.
21. Sheldrick, G.M. & Schneider, T. (1996). SHELXL: high-resolution refinement. *Methods Enzymol.* in press.
22. Luzzati, V. (1952). Traitement statistique des erreurs dans la determination des structures cristallines. *Acta Cryst*. **5**, 802–810.
23. Read, R.J. (1986). Improved Fourier coefficients for maps using phases from partial structures with errors. *Acta Cryst. A* **42**, 140–149.
24. Hsu, I.N., Delbaere, L.T.J., James, M.N.G. & Hoffman, T. (1977). Penicillopepsin from Penicillium janthinellum crystal structure at 2.8 Å and sequence homology with porcine pepsin. *Nature* **266**, 140–145.
25. Burling, F.T. & Brünger, A.T. (1994). Thermal motion and conformational disorder in protein crystal structures: comparison of multi-conformer and time-averaging models. *Isr. J. Chem*. **34**, 165–175.
26. Jiang, J.S. & Brünger, A.T. (1994). Protein hydration observed by X-ray diffraction: solvation properties of penicillopepsin and neuraminidase crystal structures. *J. Mol. Biol*. **243**, 100–115.
27. Baker, D., Bystroff, C., Fletterick, R.J. & Agard, D.A. (1993). PRISM: topologically restrained phase refinement for macromolecular crystallography. *Acta Cryst. D* **49**, 429–439.
28. Grimes, J. & Stuart, D. (1994). Use of the free R-factor as a guide in parameter optimisation for density modification. In *From First Map to Final Model*. (Bailey, S., Hubbard, R. & Waller, D., eds), pp. 67–76, SERC Daresbury Laboratory, Warrington, UK.
29. Roberts, A.L.U. & Brünger, A.T. (1995). Phase improvement by cross-validated solvent flattening. *Acta Cryst. D* **51**, 990–1002.
30. Cowtan, K.D. & Main, P. (1996). Phase combination and cross-validation in iterated density-modification procedures. *Acta Cryst. D* **52**, 43–48.
31. Rossmann, M.G. & Blow, D.M. (1962). The detection of sub-units within the asymmetric unit. *Acta Cryst*. **15**, 24–31.
32. Kleywegt, G.J., Hoier, H. & Jones, T.A. (1996). A re-evaluation of the crystal structure of chloromuconate cycloisomerase. *Acta Cryst. D* **52**, 858–863.
33. Valegård, K., Murray, J.B., Stockley, P.G., Stonehouse, N.J. & Liljas, L. (1994). Crystal structure of an RNA bacteriophage coat protein-operator complex. *Nature* **371**, 623–626.
34. Dodson, E.J., Kleywegt, G.J. & Wilson, K.S. (1996). Report of a workshop on the use of statistical validators in protein X-ray crystallography. *Acta Cryst. D* **52**, 228–234.
35. Brünger, A.T. (1990). X-PLOR: a system for crystallography and NMR, Yale University, New Haven, CT.
36. Kleywegt, G.J. & Jones, T.A. (1996). xdlMAPMAN and xdlDATAMAN - programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection datasets. *Acta Cryst. D* **52**, 826–828.
37. Diamond, R. (1971). A real-space refinement procedure for proteins. *Acta Cryst. A* **27**, 436–452.
38. Bernstein, F.C., *et al*., & Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.
39. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst. A* **47**, 110–119.